Heterogeneity
Random and fixed effects

Georgia Salanti

University of Ioannina School of Medicine
Ioannina Greece

gsalanti@cc.uoi.gr
georgia.salanti@gmail.com

Outline

- What is heterogeneity?
- Identifying heterogeneity
- Dealing with heterogeneity
- Random effects meta-analysis

- Not only for statisticians!

What is heterogeneity?

Clinical heterogeneity (clinical diversity)

- *Participants*
  - e.g. conditions under investigation, eligibility criteria for trials, geographical variation
- *Interventions*
  - e.g. intensity / dose / duration, sub-type of drug, mode of administration, experience of practitioners, nature of the control (placebo/none/standard care)
- *Outcomes*
  - e.g. definition of an event, follow-up duration, ways of measuring outcomes, cut-off points on scales

What is heterogeneity?

Methodological heterogeneity (methodological diversity)

- *Design*
  - e.g. randomised vs non-randomised,
    crossover vs parallel group vs cluster randomised,
    length
- *Conduct*
  - e.g. allocation concealment, blinding etc,
    approach to analysis, imputation methods for missing data

What is heterogeneity?

Statistical heterogeneity

- Common views
  - Variation in the results of studies
  - More variation than would be expected by chance

- In truth:
  - Variation in the *true effects* underlying the studies
  - ...which may manifest itself in more observed variation than expected by chance
  - When homogeneity does not hold (homogeneity = identical effect underlying every study)
  - May be due to different treatment effects or different biases

Some notation

- y_i : Observed effect in a study i
  - *E.g. MD, logOR etc*
- w_i : the weight of the study in the meta-analysis
  - *1/variance*
- Θ : the mean of the summary effect (meta-analysis)

Health warning for basic maths!

Identifying heterogeneity

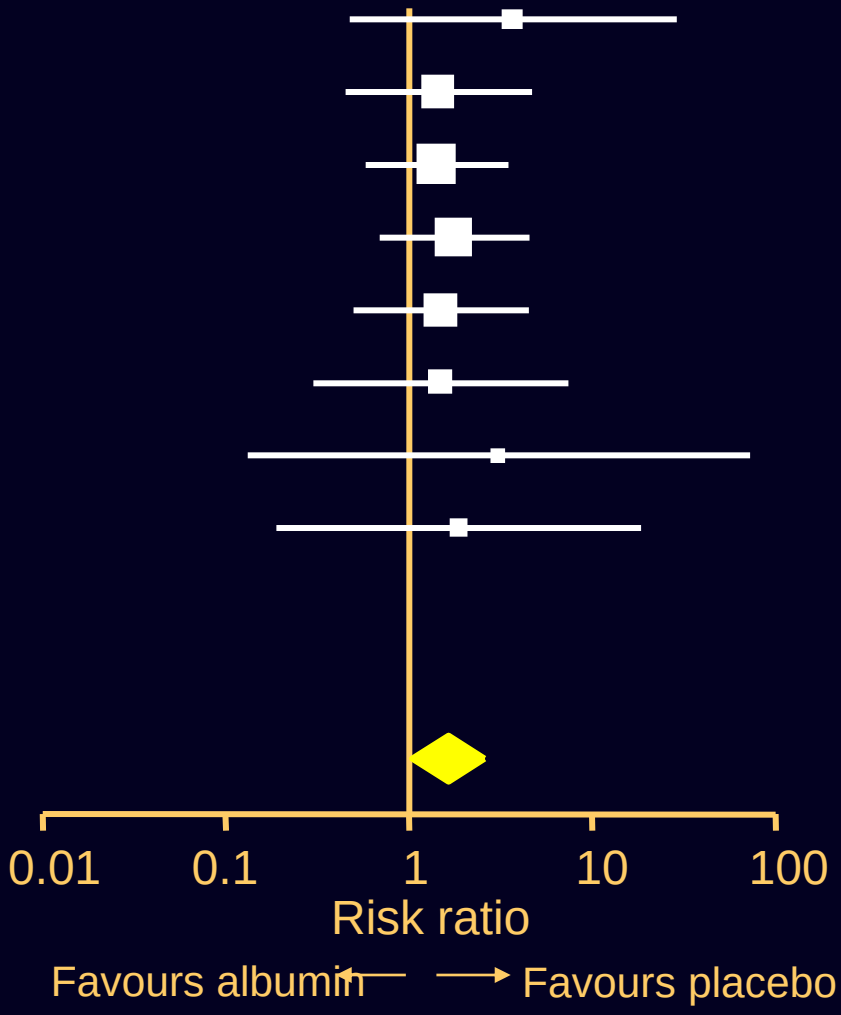
1. Visual inspection of the forest plots
2. Q test for heterogeneity
3. I^2 quantifies heterogeneity as a proportion

- You either believe in heterogeneity *a priori* or you don't

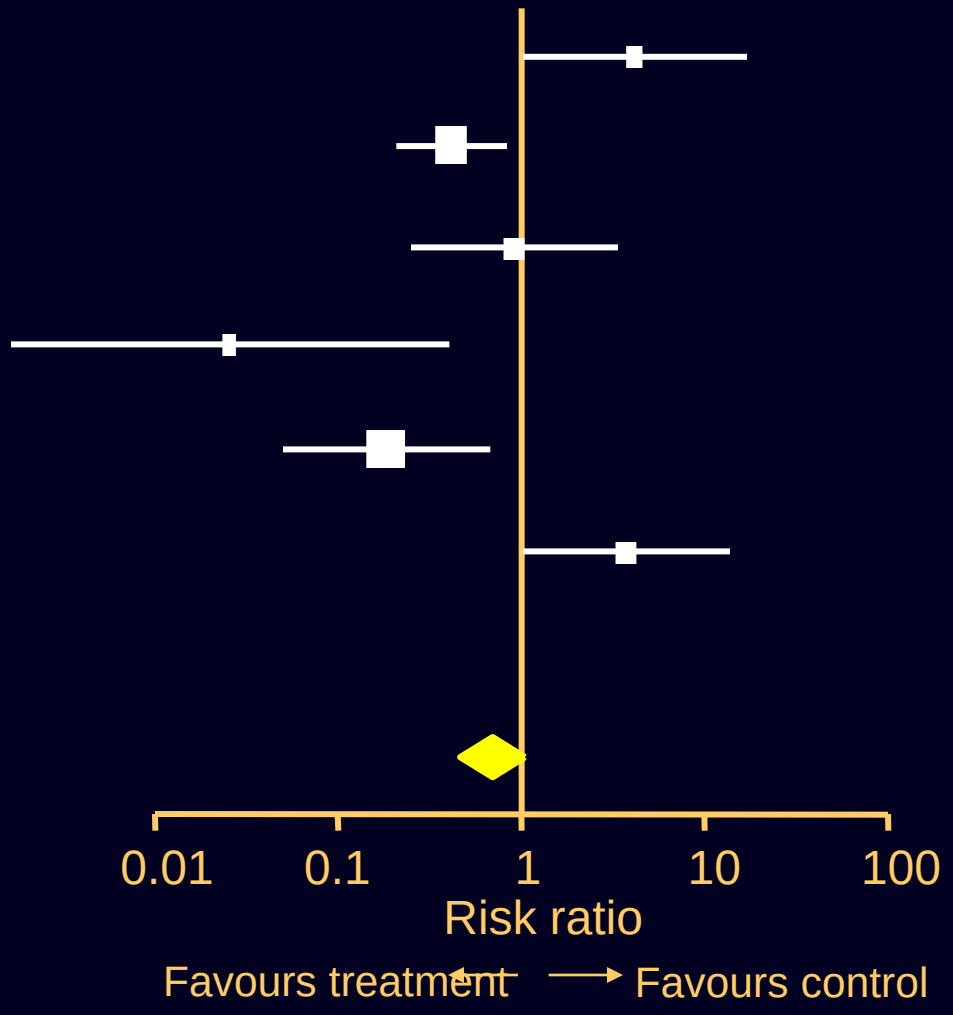
Identifying heterogeneity 1

- Eyeballing
  - a graphical inspection of the results is usually the first step
  - a lack of overlap in confidence intervals indicates heterogeneity

Albumin



Prostaglandin



Identifying heterogeneity 2

- Statistical test
  - chi-squared (χ^2) test

$$Q = \sum w_i (y_i - \theta)^2$$

  - w_i the weights, y_i the effect size in each study and θ the pooled estimate
  - has χ^2 distribution with $k - 1$ d.f. under null hypothesis of an identical effect in every study
  - *k is the number of studies in the meta-analysis*
  - rejection of H_0 suggests heterogeneity

Identifying heterogeneity 2

- Statistical test
  - Has low power since there are usually very few studies
    i.e. test is not very good at detecting heterogeneity as statistically significant when it exists
  - But, has excessive power to detect clinically unimportant heterogeneity when there are many studies

Identifying heterogeneity 3

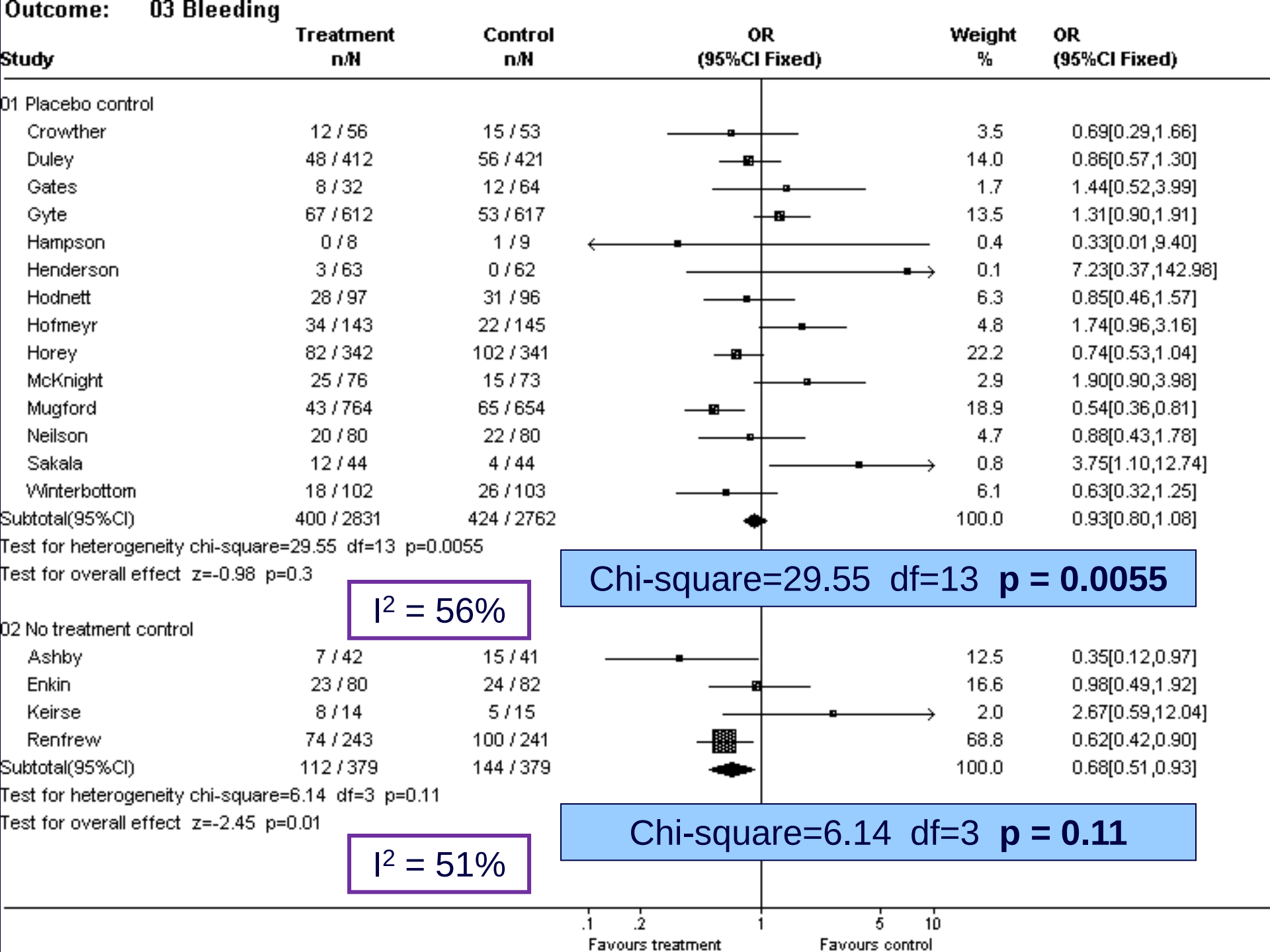
Higgins and Thompson (2002)

- Test is not asking a useful question if heterogeneity is inevitable

- Quantify heterogeneity
  - based on χ^2 statistic, Q, and its degrees of freedom.

$$I^2 = \frac{Q - k + 1}{Q} \cdot 100\%$$

describes the proportion of variability that is due to heterogeneity rather than sampling error


Outcome: 03 Bleeding
Study
Treatment n/N
Control n/N
OR (95%CI Fixed)
Weight %
OR (95%CI Fixed)
01 Placebo control
Crowther 12 / 56 15 / 53 3.5 0.69[0.29,1.66]
Duley 48 / 412 56 / 421 14.0 0.86[0.57,1.30]
Gates 8 / 32 12 / 64 1.7 1.44[0.52,3.99]
Gyte 67 / 612 53 / 617 13.5 1.31[0.90,1.91]
Hampson 0 / 8 1 / 9 0.4 0.33[0.01,9.40]
Henderson 3 / 63 0 / 62 0.1 7.23[0.37,142.98]
Hodnett 28 / 97 31 / 96 6.3 0.85[0.46,1.57]
Hofmeyr 34 / 143 22 / 145 4.8 1.74[0.96,3.16]
Horey 82 / 342 102 / 341 22.2 0.74[0.53,1.04]
McKnight 25 / 76 15 / 73 2.9 1.90[0.90,3.98]
Mugford 43 / 764 65 / 654 18.9 0.54[0.36,0.81]
Neilson 20 / 80 22 / 80 4.7 0.88[0.43,1.78]
Sakala 12 / 44 4 / 44 0.8 3.75[1.10,12.74]
Winterbottom 18 / 102 26 / 103 6.1 0.63[0.32,1.25]
Subtotal(95%CI) 400 / 2831 424 / 2762 100.0 0.93[0.80,1.08]
Test for heterogeneity chi-square=29.55 df=13 p=0.0055
Test for overall effect z=-0.98 p=0.3
I2 = 56%
Chi-square=29.55 df=13 p = 0.0055
02 No treatment control
Ashby 7 / 42 15 / 41 12.5 0.35[0.12,0.97]
Enkin 23 / 80 24 / 82 16.6 0.98[0.49,1.92]
Keirse 8 / 14 5 / 15 2.0 2.67[0.59,12.04]
Renfrew 74 / 243 100 / 241 68.8 0.62[0.42,0.90]
Subtotal(95%CI) 112 / 379 144 / 379 100.0 0.68[0.51,0.93]
Test for heterogeneity chi-square=6.14 df=3 p=0.11
Test for overall effect z=-2.45 p=0.01
I2 = 51%
Chi-square=6.14 df=3 p = 0.11
.1
.2
1
5
10
Favours treatment
Favours control

What can we do with heterogeneity?

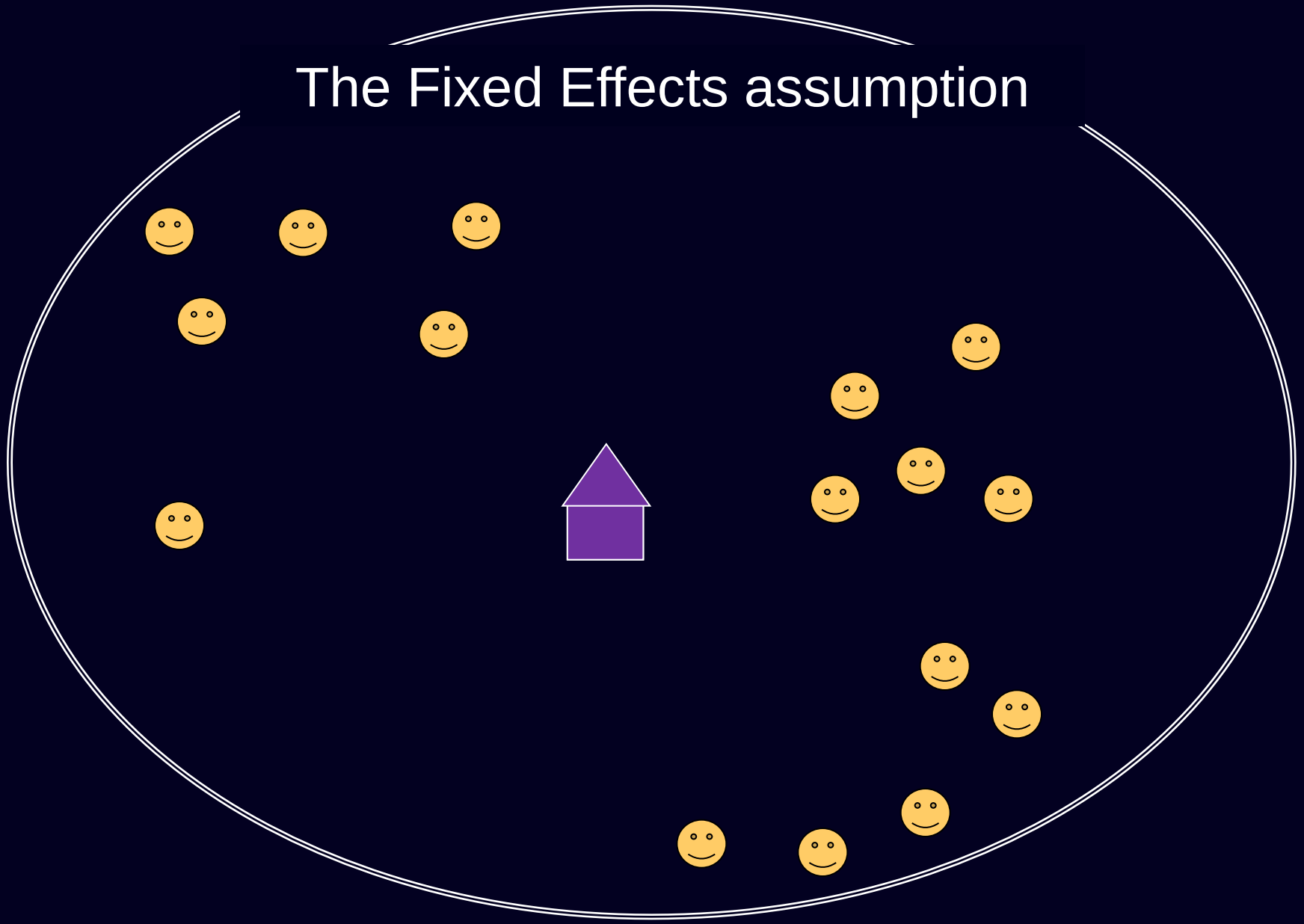
→ • Check the data	• Incorrect data extraction; unit of analysis errors (e.g. with crossover trials, cluster randomized trials, counts)
→ • Try to bypass it	• Change effect measure
• Ignore it	• Don’t do that!
• Resign to it	• Do no meta-analysis
→ • Encompass it	• Random effects meta-analysis
→ • Explore it	• Subgroup analysis Meta-regression

Random effects meta-analysis

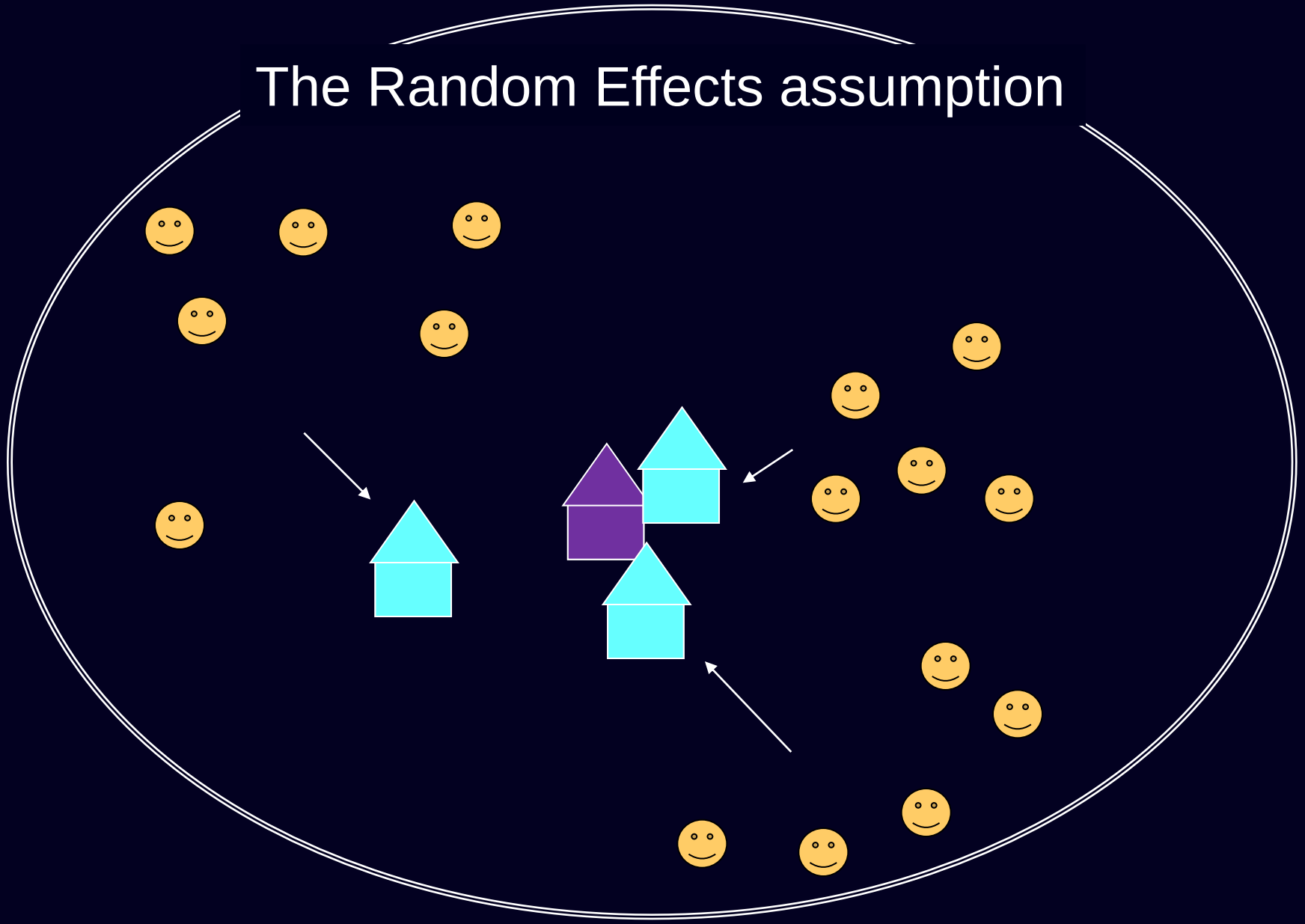
- Heterogeneity suggests that the studies have important underlying differences
- We can allow the true effects underlying the studies to differ
- We assume the true effects underlying the studies follow a distribution
  - conventionally a normal distribution
- It turns out that we can use a simple adaptation of the inverse-variance weighted average

DerSimonian and Laird (1986)

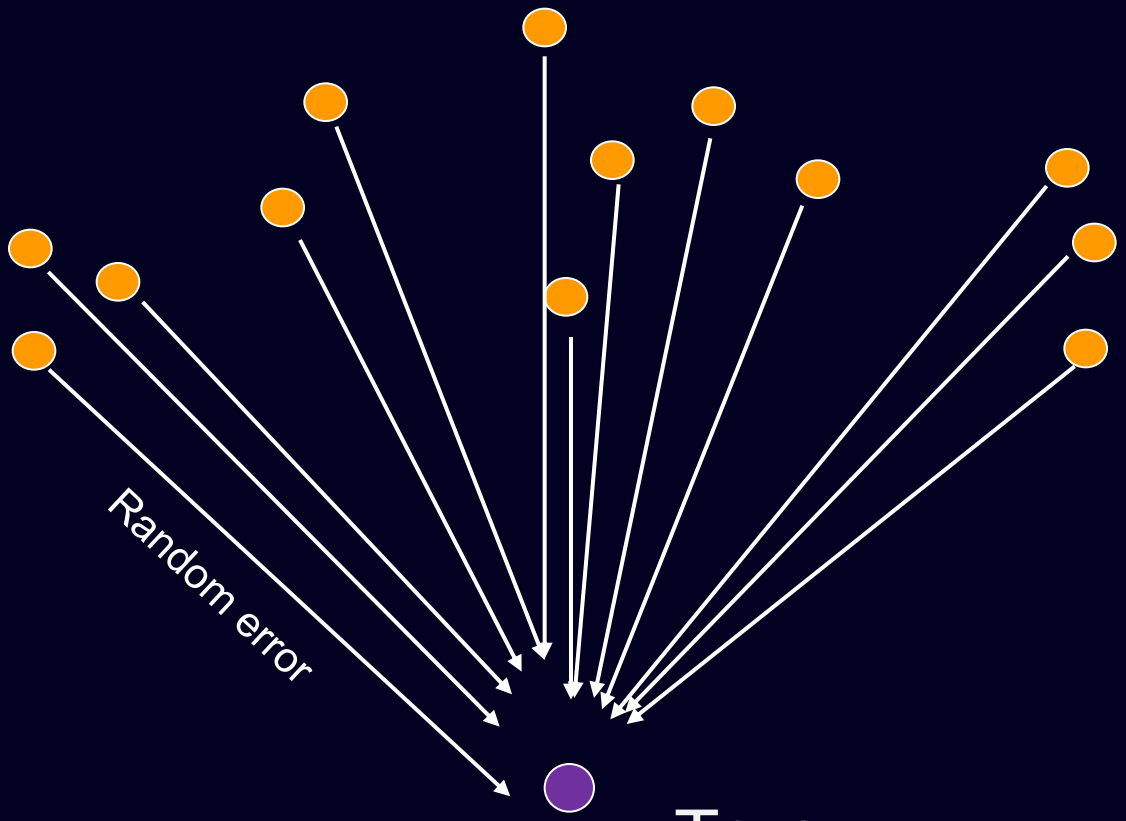
The Fixed Effects assumption



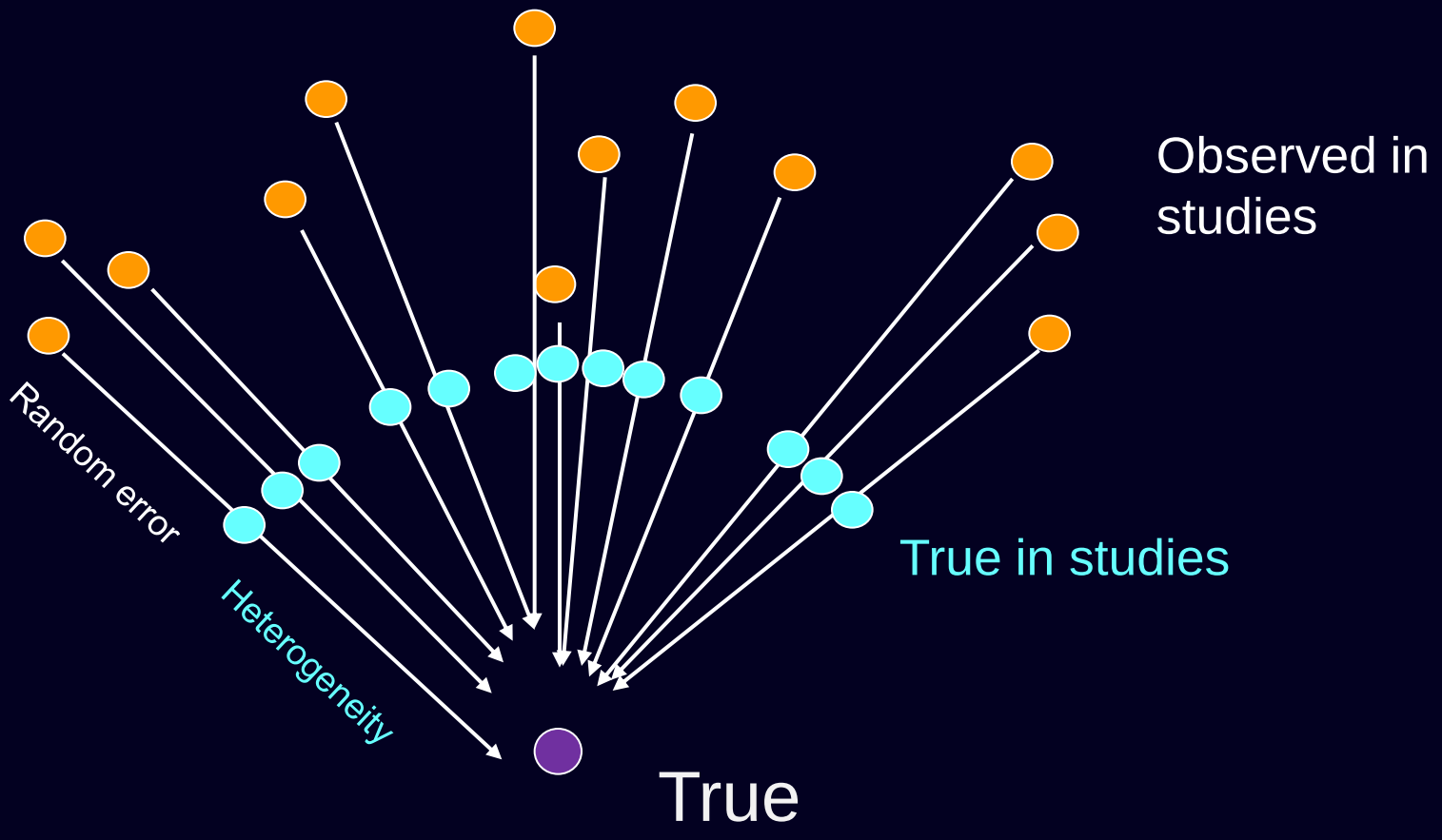
The Random Effects assumption



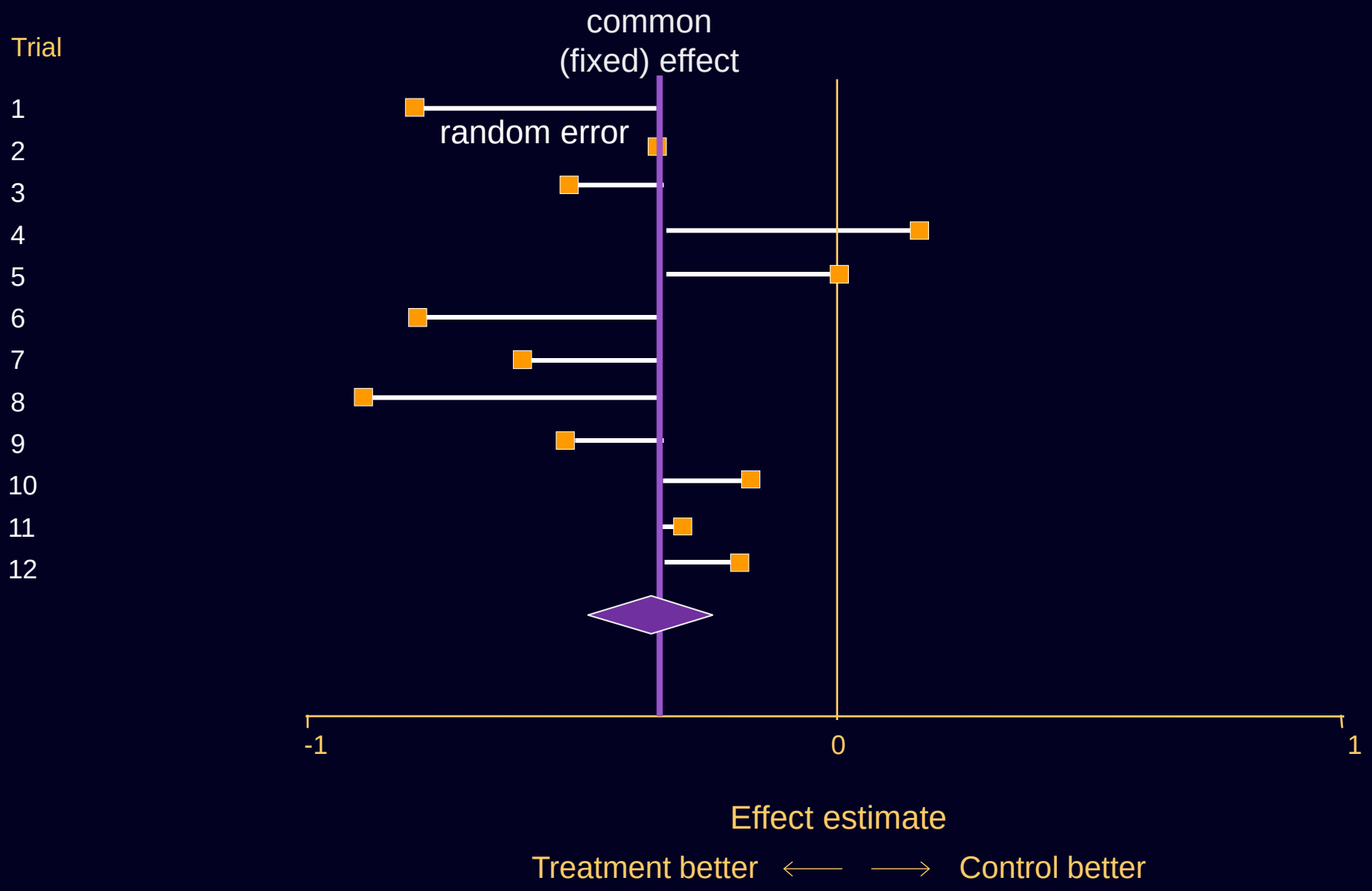
The Fixed Effects assumption



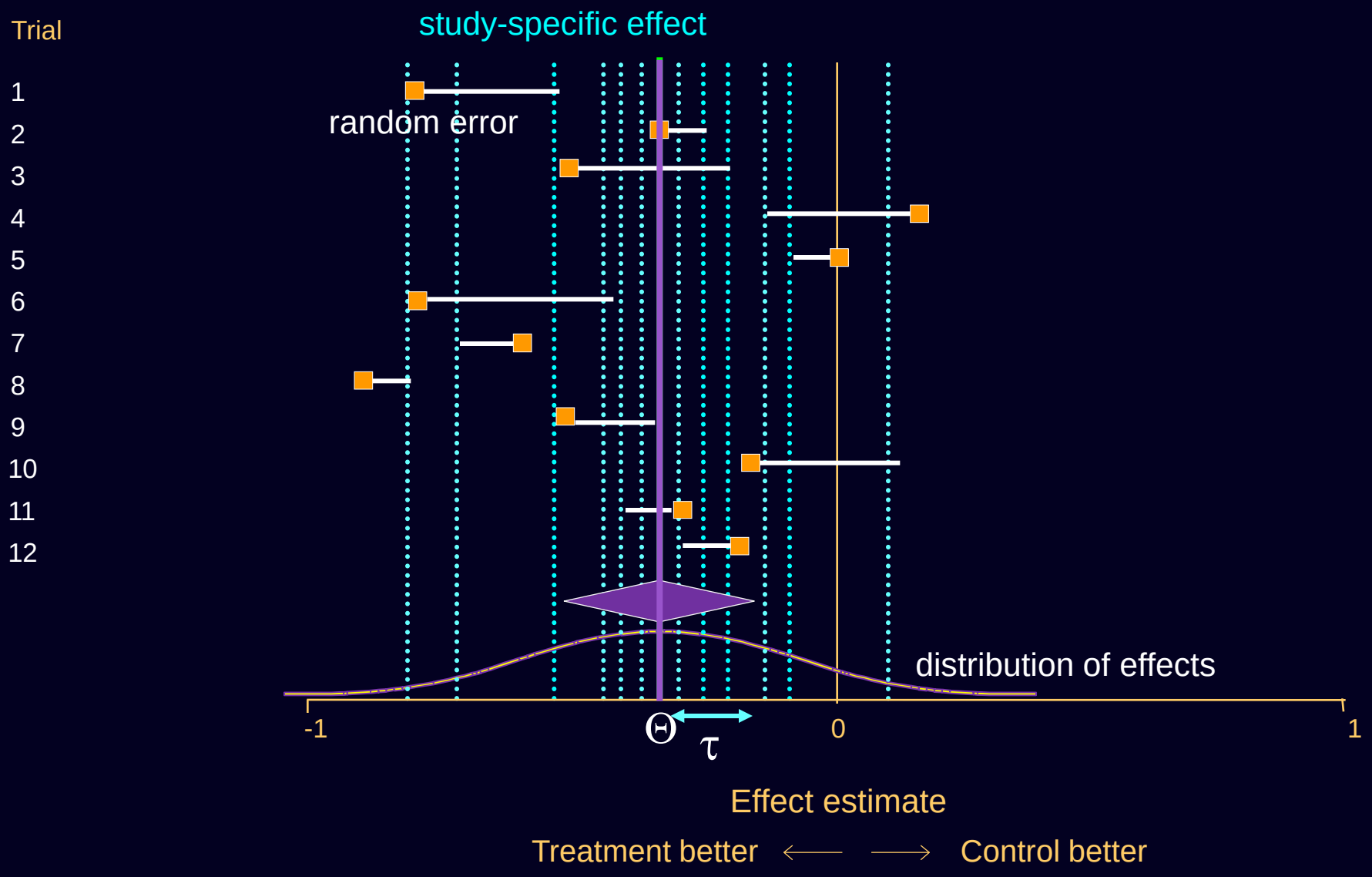
The Random Effects assumption



Fixed effect meta-analysis



Random effects meta-analysis



Random effects: estimation

- It is a ‘simple’ extension of the inverse variance method, by taking into account the variance of the random effects τ^2

Three steps:

1. Calculate τ^2 (also called the heterogeneity parameter)
2. Re-define the weights w_i^*
3. Calculate the pooled estimates and its variance using the weights w_i^*

Step 1: Estimate τ^2

- estimate the heterogeneity variance τ^2 from the test Q (method of moments) :

$$\tau^2 = \frac{Q - (k-1)}{\sum w_i - \left(\sum w_i^2\right) / \sum w_i}$$

- We set $\tau^2 = 0$ if $Q < (k-1)$

Step 2: re-define the weights

- We incorporate the heterogeneity parameter in the study weights:

$$w_i^* = \frac{1}{v_i + \hat{\tau}^2}$$

Note that the weights are smaller than previously

where v_i is the variance in study i

Step 3: Calculate the pooled estimate

Using the inverse variance method with the new weights

$$\theta = \frac{\sum w_i^* y_i}{\sum w_i^*}$$

$$SE(\theta) = \frac{1}{\sqrt{\sum w_i^*}}$$

Example (organised inpatient rehabilitation review)

calculations

Study	OR	Ln OR y_i	Var v_i	Weight w_i	$w_i y_i$	$w_i y_i^2$	w_i^2
Cameron 1993	0.98	–0.02	0.10	9.7	–0.19	0.004	94.4
Fordham 1986	1.36	0.31	0.26	3.9	1.2	0.37	15.2
Galvard 1995	1.28	0.25	0.06	16.4	4.1	1.01	269.3
Gilchrist 1988	0.75	–0.29	0.14	7.3	–2.1	0.62	53.5
Kennie 1988	0.45	–0.79	0.21	4.9	–3.8	3.02	23.6
Total				**42.2**	**–0.87**	**5.01**	**455.9**

$$Q = 5.01 - \frac{(-0.87)^2}{42.2} = 5.00$$

$$\tau^2 = \frac{5.00 - 4}{42.2 - 455.9/42.2} = 0.0319$$

Example *continued*

Study	OR	Ln OR y_i	Var v_i	Weight w^*_i	$w^*_i y_i$
Cameron 1993	0.98	–0.02	0.10	7.6	-0.2
Fordham 1986	1.36	0.31	0.26	3.4	1.1
Galvard 1995	1.28	0.25	0.06	10.9	2.7
Gilchrist 1988	0.75	–0.29	0.14	5.8	-1.7
Kennie 1988	0.45	–0.79	0.21	4.1	-3.3
Total				**31.8**	**-1.3**

Meta-analysis result (for logOR):

θ = – 0.04

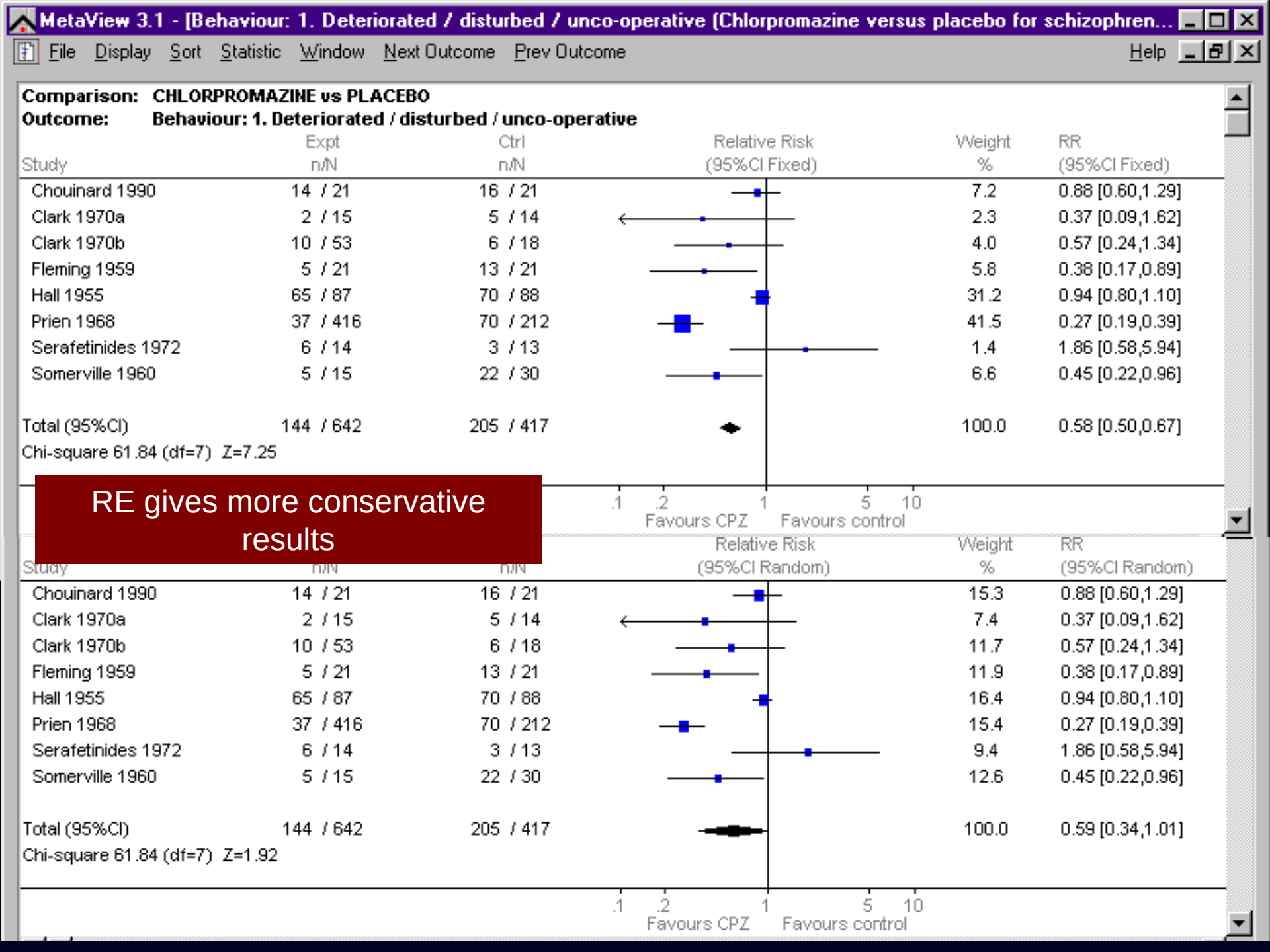
SE(θ) = 0.177

95%CI: – 0.39 to 0.30

- Back on odds ratio scale:
  - pooled odds ratio = exp{– 0.045} = 0.96
  - 95% confidence interval from 0.68 to 1.35

- *c.f. fixed effect analysis*
  - *pooled odds ratio = exp{– 0.02} = 0.98*
  - *95% confidence interval from 0.72 to 1.32*

Random effects give wider confidence intervals!


MetaView 3.1 - [Behaviour: 1. Deteriorated / disturbed / unco-operative (Chlorpromazine versus placebo for schizophren...
File Display Sort Statistic Window Next Outcome Prev Outcome Help
Comparison: CHLORPROMAZINE vs PLACEBO
Outcome: Behaviour: 1. Deteriorated / disturbed / unco-operative
Study
Expt n/N
Ctrl n/N
Relative Risk (95%CI Fixed)
Weight %
RR (95%CI Fixed)
Chouinard 1990 14 / 21 16 / 21 7.2 0.88 [0.60,1.29]
Clark 1970a 2 / 15 5 / 14 2.3 0.37 [0.09,1.62]
Clark 1970b 10 / 53 6 / 18 4.0 0.57 [0.24,1.34]
Fleming 1959 5 / 21 13 / 21 5.8 0.38 [0.17,0.89]
Hall 1955 65 / 87 70 / 88 31.2 0.94 [0.80,1.10]
Prien 1968 37 / 416 70 / 212 41.5 0.27 [0.19,0.39]
Serafetinides 1972 6 / 14 3 / 13 1.4 1.86 [0.58,5.94]
Somerville 1960 5 / 15 22 / 30 6.6 0.45 [0.22,0.96]
Total (95%CI) 144 / 642 205 / 417 100.0 0.58 [0.50,0.67]
Chi-square 61.84 (df=7) Z=7.25
.1 .2 1 5 10
Favours CPZ Favours control
RE gives more conservative results
Study
n/N
n/N
Relative Risk (95%CI Random)
Weight %
RR (95%CI Random)
Chouinard 1990 14 / 21 16 / 21 15.3 0.88 [0.60,1.29]
Clark 1970a 2 / 15 5 / 14 7.4 0.37 [0.09,1.62]
Clark 1970b 10 / 53 6 / 18 11.7 0.57 [0.24,1.34]
Fleming 1959 5 / 21 13 / 21 11.9 0.38 [0.17,0.89]
Hall 1955 65 / 87 70 / 88 16.4 0.94 [0.80,1.10]
Prien 1968 37 / 416 70 / 212 15.4 0.27 [0.19,0.39]
Serafetinides 1972 6 / 14 3 / 13 9.4 1.86 [0.58,5.94]
Somerville 1960 5 / 15 22 / 30 12.6 0.45 [0.22,0.96]
Total (95%CI) 144 / 642 205 / 417 100.0 0.59 [0.34,1.01]
Chi-square 61.84 (df=7) Z=1.92
.1 .2 1 5 10
Favours CPZ Favours control

Aside: a note on RevMan

- For binary data, the Q statistic in RevMan is calculated using the Mantel-Haenszel estimate of θ rather than the usual inverse-variance weighted average.

- Thus results can be slightly different

- STATA's *metan* uses the same Mantel-Haenszel variation by default (options *fixed* and *random*).
- To apply the method described here, use *fixedi* and *randomi*)

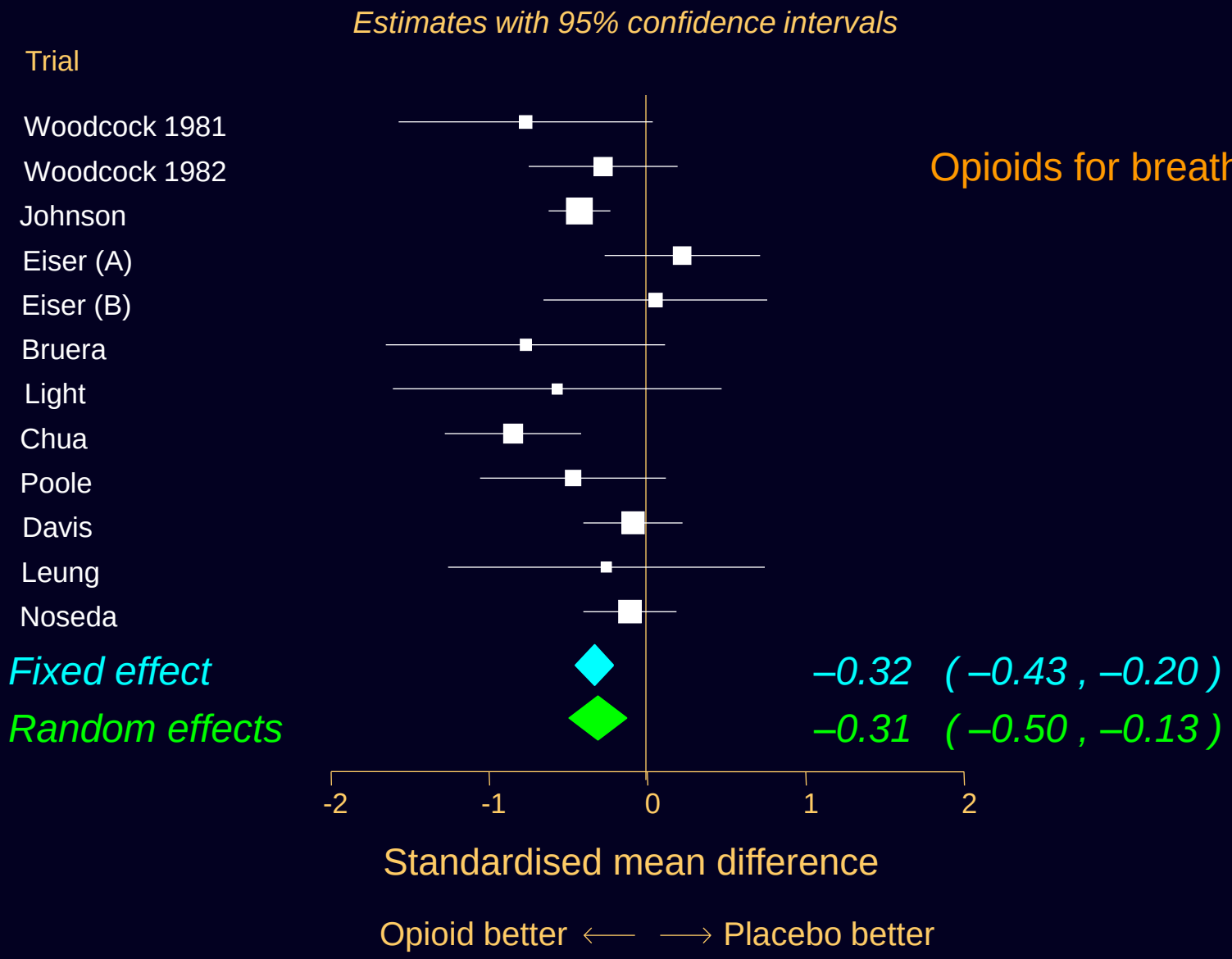
What about fixed vs random effects?

- Fixed effect model often unrealistic
- Random effects model difficult to justify
- Random effects analysis may give spurious results when effect size depends on precision
  - (gives relatively more weight to smaller studies)
  - Important because
    - Smaller studies may be of lower quality (hence biased)
    - Publication bias may result in missing smaller studies
- Fixed versus random effects *is an ongoing debate*

Comparison of fixed and random effects meta-analyses

- Fixed and random effects inverse-variance meta-analyses may
  - be identical (when $\tau^2 = 0$)
  - give similar point estimate, different confidence intervals

Fixed versus random effects: Slightly different results



Comparison of fixed and random effects meta-analyses

- Fixed and random effects inverse-variance meta-analyses may
  - be identical (when $\tau^2 = 0$)
  - give similar point estimate, different confidence intervals
  - give different point estimates, different confidence intervals

- The last happens when there is funnel plot asymmetry

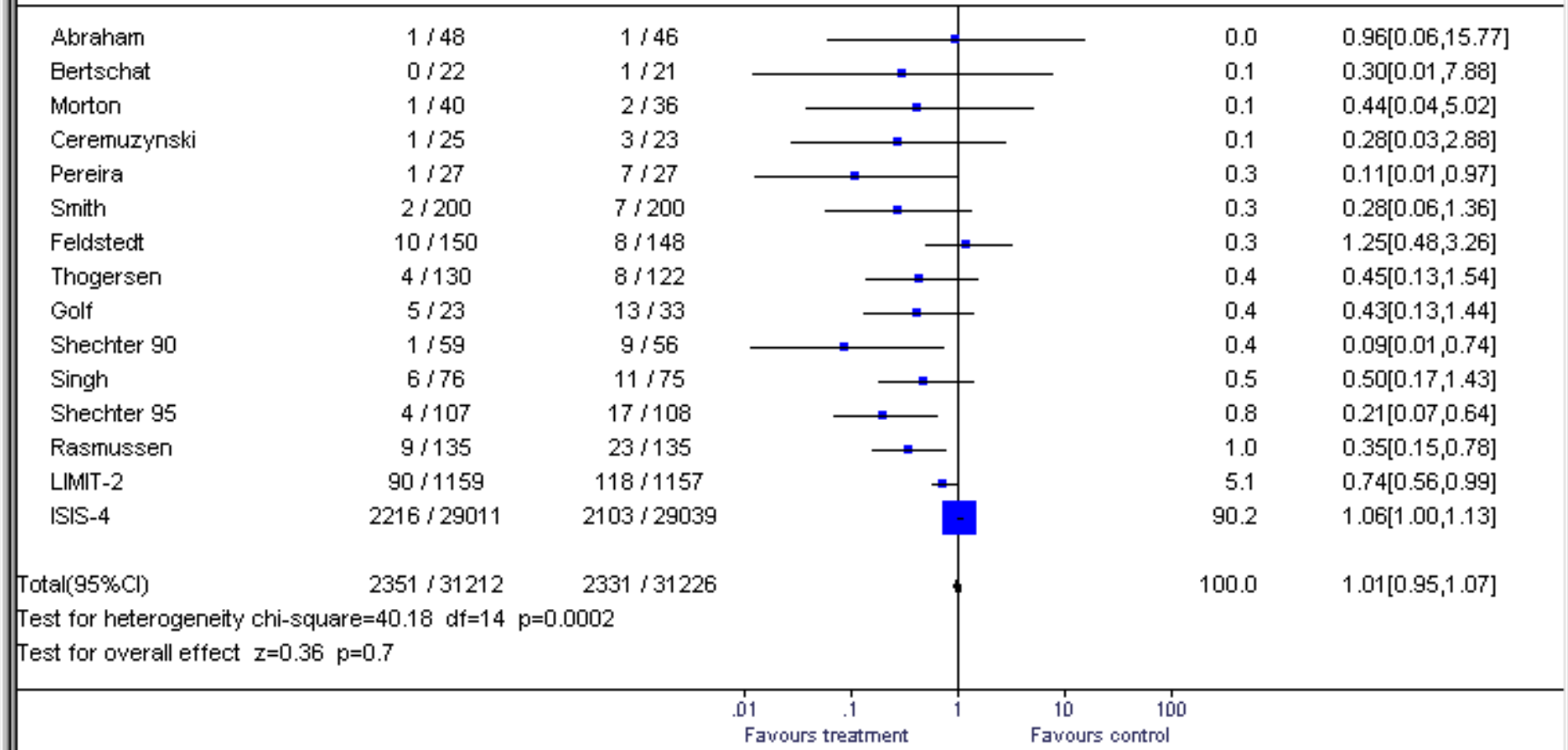
METAVIEW 4.1 [Review: Magnesium for acute AI]

File Display Sort Statistics Previous Outcome Next Outcome Help

Comparison: 13 Magnesium vs placebo
Outcome: 03 Mortality

Study	Treatment n/N	Control n/N	OR (95%CI Fixed)	Weight %	OR (95%CI Fixed)
Abraham	1 / 48	1 / 46		0.0	0.96[0.06,15.77]
Bertschat	0 / 22	1 / 21		0.1	0.30[0.01,7.88]
Morton	1 / 40	2 / 36		0.1	0.44[0.04,5.02]
Ceremuzynski	1 / 25	3 / 23		0.1	0.28[0.03,2.88]
Pereira	1 / 27	7 / 27		0.3	0.11[0.01,0.97]
Smith	2 / 200	7 / 200		0.3	0.28[0.06,1.36]
Feldstedt	10 / 150	8 / 148		0.3	1.25[0.48,3.26]
Thogersen	4 / 130	8 / 122		0.4	0.45[0.13,1.54]
Golf	5 / 23	13 / 33		0.4	0.43[0.13,1.44]
Shechter 90	1 / 59	9 / 56		0.4	0.09[0.01,0.74]
Singh	6 / 76	11 / 75		0.5	0.50[0.17,1.43]
Shechter 95	4 / 107	17 / 108		0.8	0.21[0.07,0.64]
Rasmussen	9 / 135	23 / 135		1.0	0.35[0.15,0.78]
LIMIT-2	90 / 1159	118 / 1157		5.1	0.74[0.56,0.99]
ISIS-4	2216 / 29011	2103 / 29039		90.2	1.06[1.00,1.13]
Total(95%CI)	2351 / 31212	2331 / 31226		100.0	1.01[0.95,1.07]

Test for heterogeneity chi-square=40.18 df=14 p=0.0002
Test for overall effect z=0.36 p=0.7



METAVIEW 4.1 [Review: Magnesium for acute AI]

File Display Sort Statistics Previous Outcome Next Outcome Help

Comparison: 13 Magnesium vs placebo
Outcome: 03 Mortality

Study	Treatment n/N	Control n/N	OR (95%CI Random)	Weight %	OR (95%CI Random)
Abraham	1 / 48	1 / 46		1.6	0.96[0.06,15.77]
Bertschat	0 / 22	1 / 21		1.2	0.30[0.01,7.88]
Morton	1 / 40	2 / 36		2.1	0.44[0.04,5.02]
Ceremuzynski	1 / 25	3 / 23		2.3	0.28[0.03,2.88]
Pereira	1 / 27	7 / 27		2.6	0.11[0.01,0.97]
Smith	2 / 200	7 / 200		4.3	0.28[0.06,1.36]
Feldstedt	10 / 150	8 / 148		8.6	1.25[0.48,3.26]
Thogersen	4 / 130	8 / 122		6.3	0.45[0.13,1.54]
Golf	5 / 23	13 / 33		6.4	0.43[0.13,1.44]
Shechter 90	1 / 59	9 / 56		2.7	0.09[0.01,0.74]
Singh	6 / 76	11 / 75		7.7	0.50[0.17,1.43]
Shechter 95	4 / 107	17 / 108		7.1	0.21[0.07,0.64]
Rasmussen	9 / 135	23 / 135		10.2	0.35[0.15,0.78]
LIMIT-2	90 / 1159	118 / 1157		17.5	0.74[0.56,0.99]
ISIS-4	2216 / 29011	2103 / 29039		19.4	1.06[1.00,1.13]
Total(95%CI)	2351 / 31212	2331 / 31226		100.0	0.53[0.36,0.77]

Test for heterogeneity chi-square=40.18 df=14 p=0.0002
Test for overall effect z=-3.34 p=0.0008

.01 .1 1
Favours treatment

Was 0%

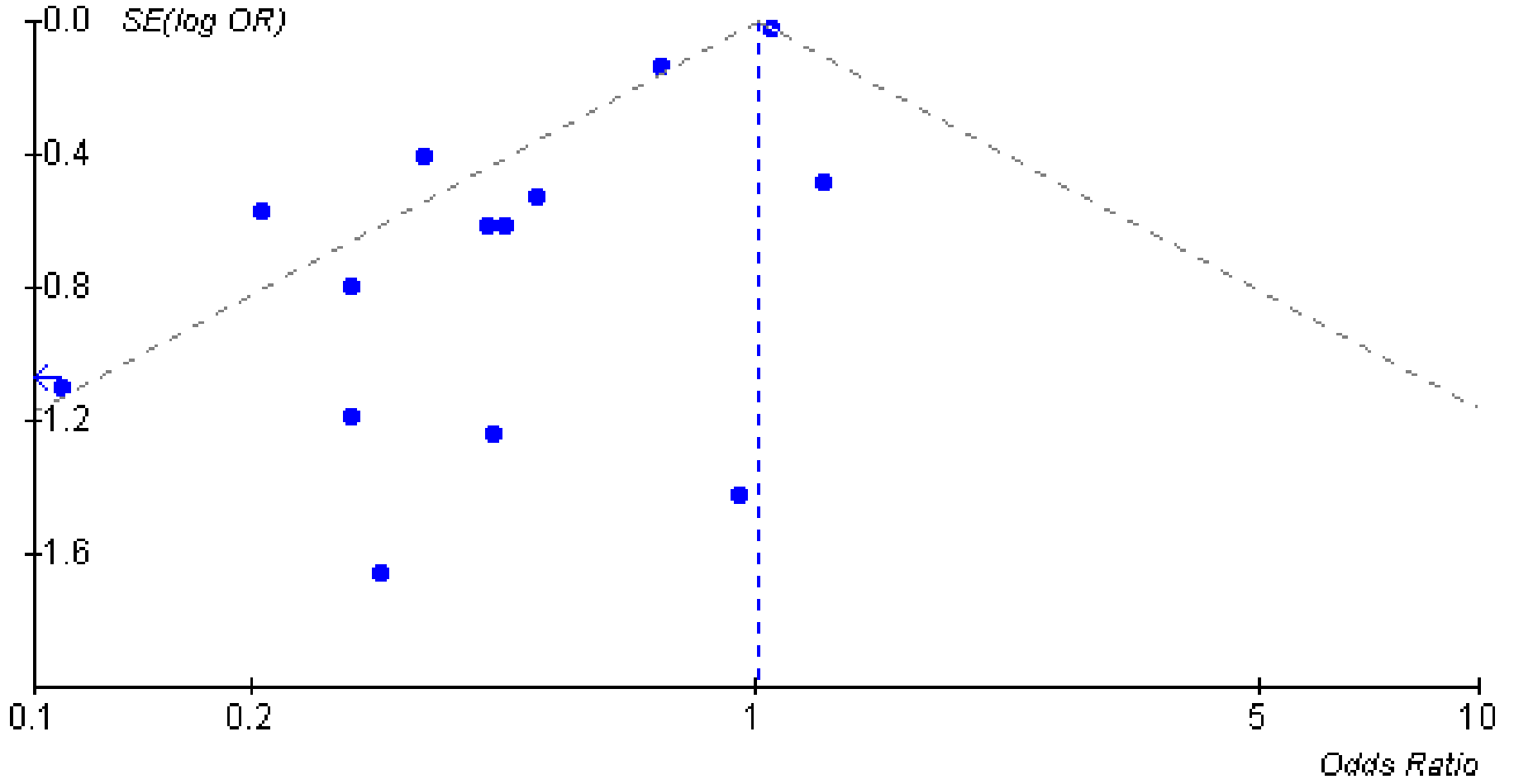
Was 90%

RE gives less 'contrasted' weights between big and small studies

Funnel plot

File View Help

Review: Demonstration
Comparison: 13 Magnesium vs placebo
Outcome: 03 Mortality

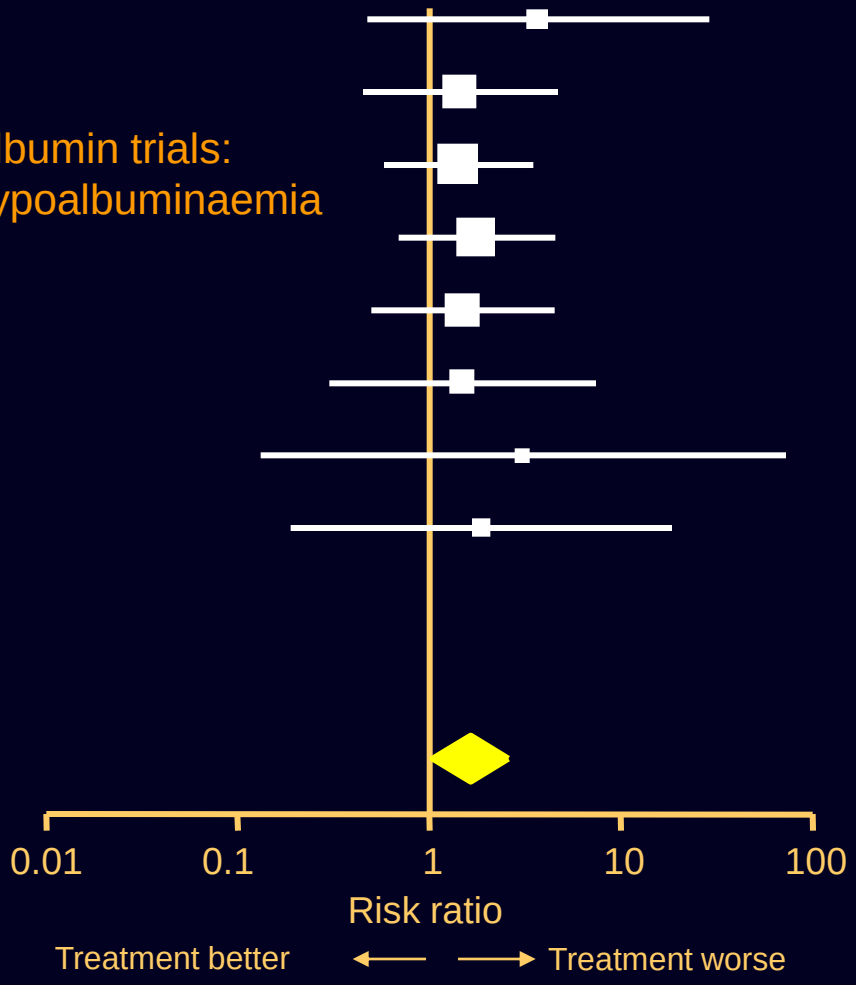


Interpreting random-effects meta-analysis

- Random-effects meta-analysis suitable for unexplained heterogeneity
  - May be unsuitable if important covariates are available
- Conventionally, inference is focused on the mean of the distribution (θ)
  - i.e. we report mean and 95% from a meta-analysis
- This may be misleading...
- Better look at the predictive interval

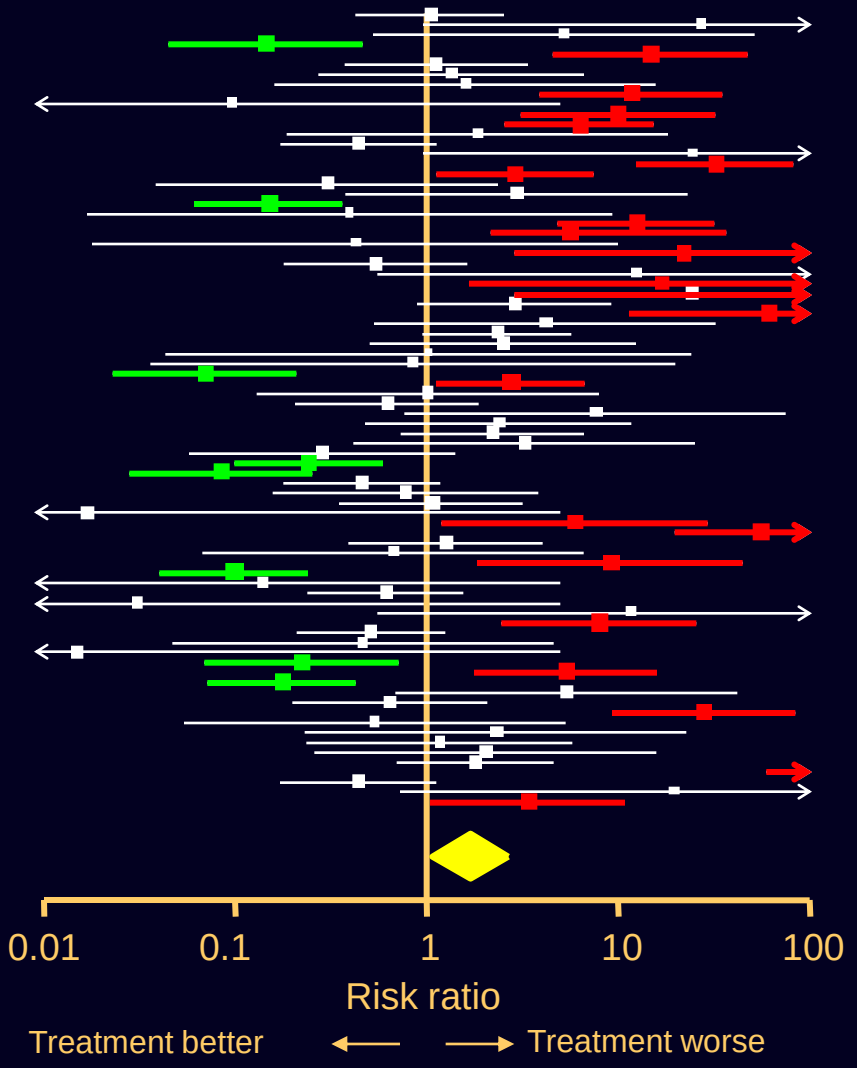
Estimates with 95% confidence intervals

Albumin trials:
hypoalbuminaemia



Random effects meta-analysis:
1.64 (1.04 , 2.58) P = 0.03

Estimates with 95% confidence intervals



Random effects meta-analysis:
1.64 (1.04 , 2.58) P = 0.03

Interpreting random-effects meta-analysis

- Random-effects meta-analysis suitable for unexplained heterogeneity
  - May be unsuitable if important covariates are available
- Conventionally, inference is focused on the mean of the distribution (θ)
  - i.e. we report mean and 95% from a meta-analysis
- This may be misleading...
- Better look at the predictive interval
- *But this is not currently implemented in software...*
- e.g. mean $\pm$ 1.96 $\times$ τ

  or a predictive distribution for the effect in the next study

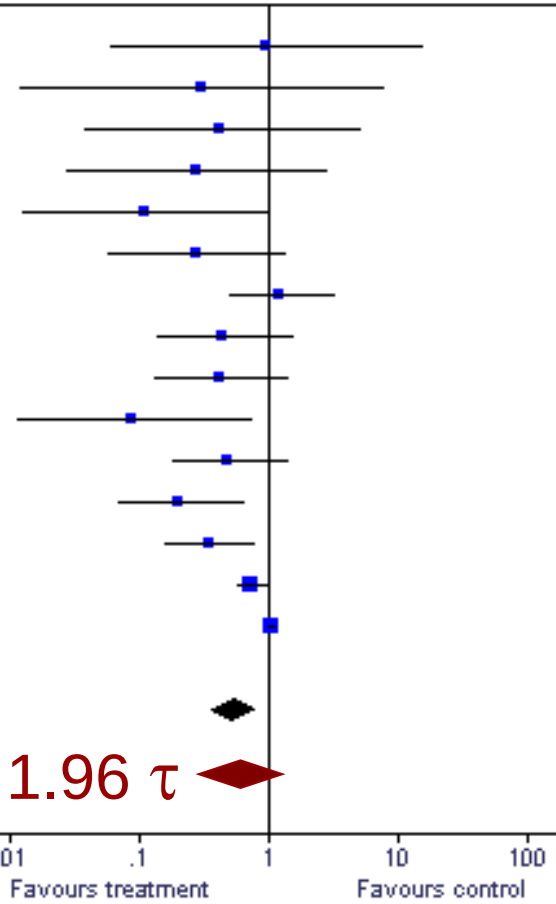
METAVIEW 4.1 [Review: Magnesium for acute AI]

File Display Sort Statistics Previous Outcome Next Outcome Help

Comparison: 13 Magnesium vs placebo
Outcome: 03 Mortality

Study	Treatment n/N	Control n/N	OR (95%CI Random)	Weight %	OR (95%CI Random)
Abraham	1 / 48	1 / 46		1.6	0.96[0.06,15.77]
Bertschat	0 / 22	1 / 21		1.2	0.30[0.01,7.88]
Morton	1 / 40	2 / 36		2.1	0.44[0.04,5.02]
Ceremuzynski	1 / 25	3 / 23		2.3	0.28[0.03,2.88]
Pereira	1 / 27	7 / 27		2.6	0.11[0.01,0.97]
Smith	2 / 200	7 / 200		4.3	0.28[0.06,1.36]
Feldstedt	10 / 150	8 / 148		8.6	1.25[0.48,3.26]
Thogersen	4 / 130	8 / 122		6.3	0.45[0.13,1.54]
Golf	5 / 23	13 / 33		6.4	0.43[0.13,1.44]
Shechter 90	1 / 59	9 / 56		2.7	0.09[0.01,0.74]
Singh	6 / 76	11 / 75		7.7	0.50[0.17,1.43]
Shechter 95	4 / 107	17 / 108		7.1	0.21[0.07,0.64]
Rasmussen	9 / 135	23 / 135		10.2	0.35[0.15,0.78]
LIMIT-2	90 / 1159	118 / 1157		17.5	0.74[0.56,0.99]
ISIS-4	2216 / 29011	2103 / 29039		19.4	1.06[1.00,1.13]
Total(95%CI)	2351 / 31212	2331 / 31226		100.0	0.53[0.36,0.77]

Test for heterogeneity chi-square=40.18 df=14 p=0.0002
Test for overall effect z=-3.34 p=0.0008



What not to do!

- Fixed or random effects meta-analysis should be specified *a priori* if possible and not on the basis of the Q test

What to do:
Think about the question you asked, the included studies etc: do you expect them to be very diverse?

You can apply and present both fixed and random effects

Exploring heterogeneity

In which types of trials does the intervention work best?

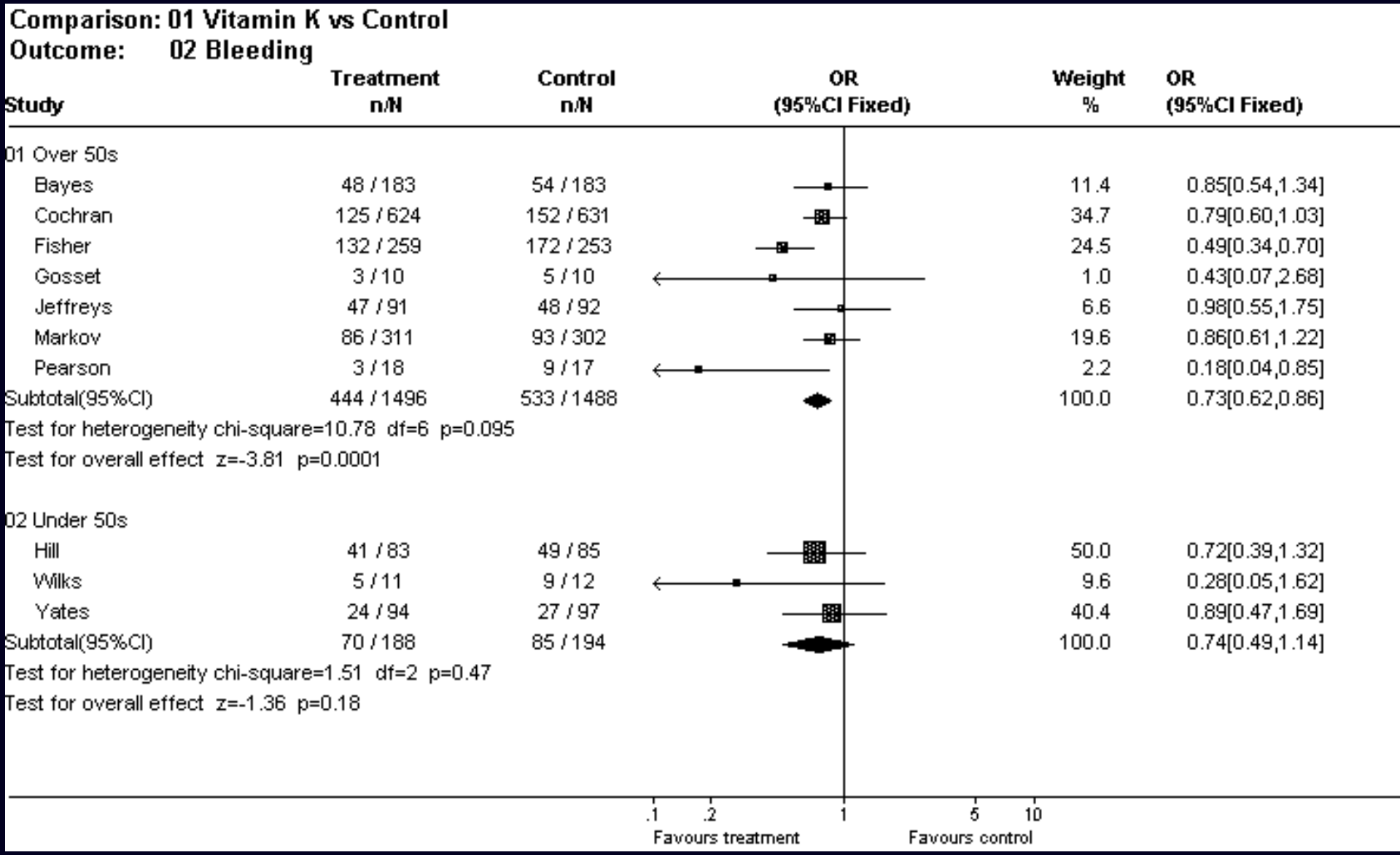
- Characteristics of studies may be associated with the size of treatment effect

- For example,
  - adequate allocation concealment
  - age group of patients
  - setting of study
  - dose of drug

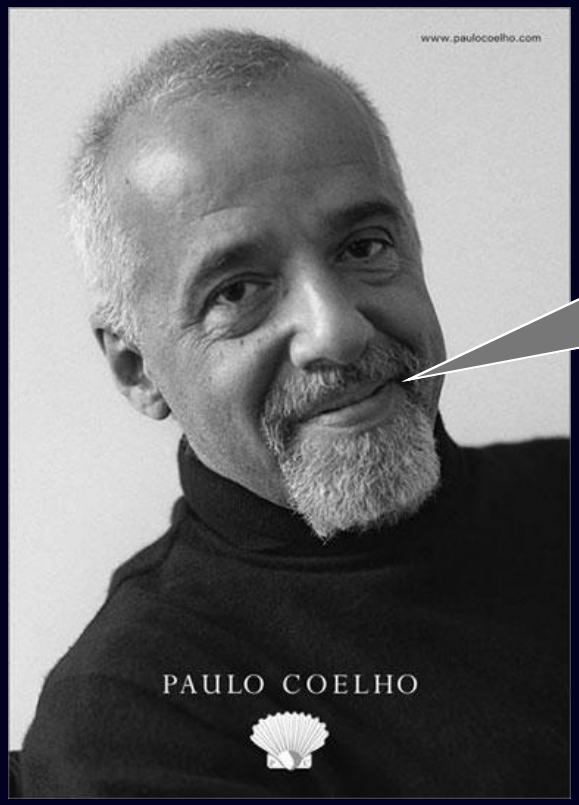
- For discrete characteristics, can use subgroup analyses

- For discrete or continuous characteristics, can use meta-regression

Vitamin K for Bleeding: *Subgroup analysis*



Subgroup analysis: dangers



Identify a priori the variables to use in subgroup analysis!

Or: *"If you look really hard in all possible subgroups, using any possible variable, then, you will find something significant"*

Selecting subgroups

- Specify characteristics in advance
- Select a small number of characteristics
- Ensure there is scientific rationale for investigating the characteristics
  - beware ‘prognostic factors’
- Make sure the effect of a characteristic can be identified
  - does it differentiate studies?
- Think about whether the characteristic is closely related to another characteristic

Meta-regression

- Relate size of treatment effect to numerical characteristic(s) of the trials
- Characteristics can be continuous or categorical
- Categorical characteristics enable a formal (but not a safer) approach to subgroup analyses
- The relationship is like a traditional regression

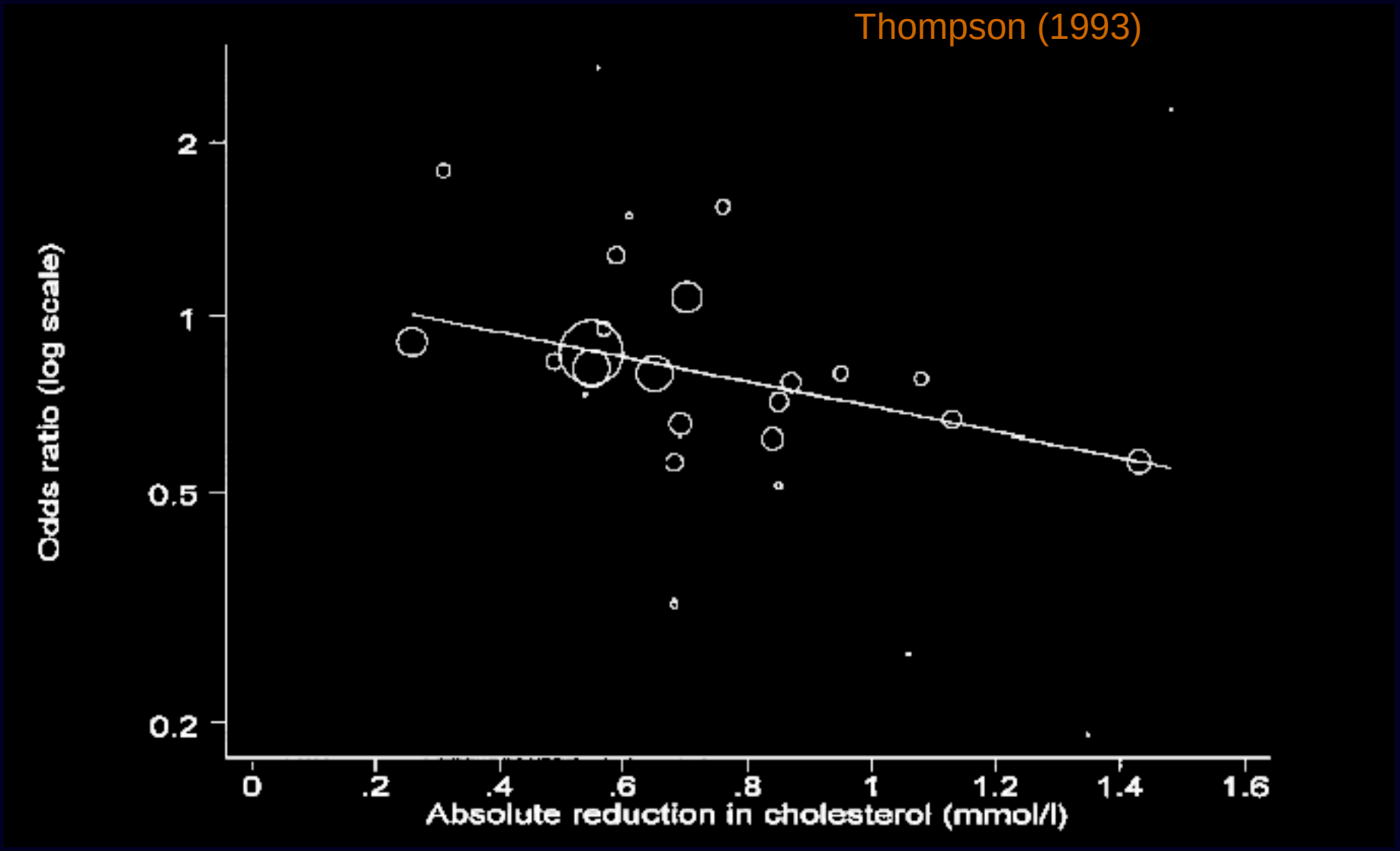
$$y_i \sim N(\theta + \beta x_i, \text{var}(y_i))$$

allowing for heteroscedasticity

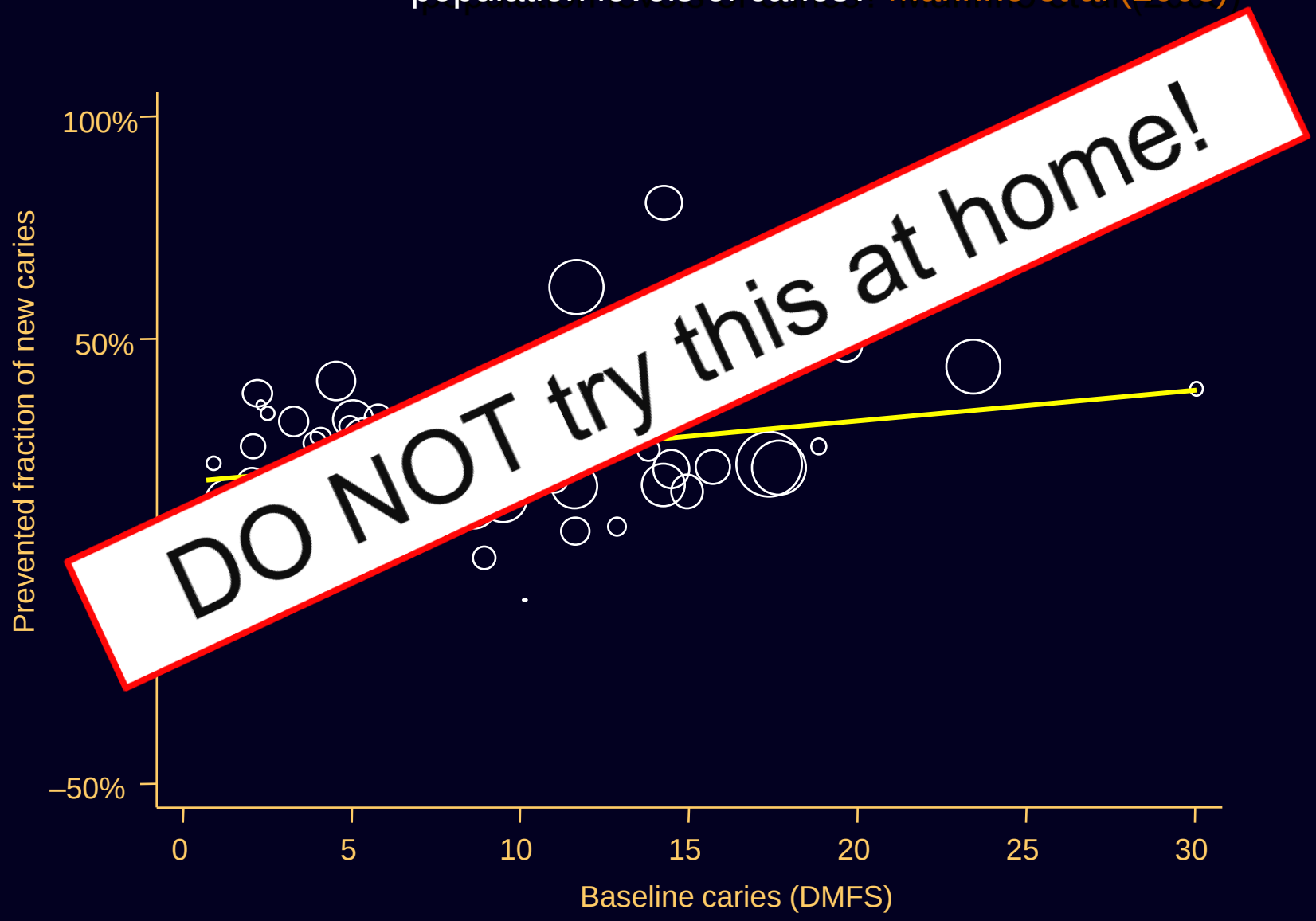
- We estimate the slope, β

Estimated odds ratios for IHD events according to extent of serum cholesterol reduction

Thompson (1993)



Does effectiveness of toothpaste depend on baseline population levels of caries? Marinho et al (2003)



Methods available in RevMan 4.2

- For each meta-analysis, or subgroup of studies:
  - Estimate of overall effect with CI (fixed effect model)
  - Estimate of mean effect with CI (random effects model)
  - Test for heterogeneity, with P value
  - I^2 measure of inconsistency
  - τ^2 heterogeneity variance
  - Test for subgroup differences

Methods not available in RevMan

- Meta-regression
- Random effects meta-analysis methods that account for the fact the τ^2 is estimated
- Predictive intervals

References

- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986; **7**: 177-188
- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539-1558
- Higgins JPT, Thompson SG. Controlling the risk of spurious results from meta-regression. *Stat Med* 2004 (in press)
- Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Stat Methods Med Res* 1993; **2**: 173-192
- Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* 1994; **309**: 1351-1355
- Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; **21**: 1559-1574
- Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999; **18**: 2693-2708